

## Probabilistic photovoltaic forecasting combining heterogeneous sources of input data for multiple time-frames

Thomas Carriere<sup>1,2</sup>, Christophe Vernay<sup>3</sup>, Sebastien Pitaval<sup>2,3</sup>, François-Pascal Neirac<sup>1</sup>,  
George Kariniotakis<sup>1</sup>

<sup>1</sup> MINES ParisTech, PSL University, Centre PERSEE, Sophia-Antipolis, France

<sup>2</sup> Third Step Energy, Sophia-Antipolis, France

<sup>3</sup> Solaïs, Sophia-Antipolis, France

The valorization of photovoltaic (PV) energy generation involves several decision making processes at different times with different objectives. For example, a PV power plant coupled with a Battery Energy Storage System (BESS) has to provide bids in the day-ahead electricity market, but can also provide ancillary services. On the delivery day, it can also participate in intra-day trading sessions, and must decide which quantity to charge or discharge from the BESS in real-time. These successive decision-making processes all require forecasts of the energy production level for different forecast horizons.

However, the models and the inputs used for the different forecast horizons are often different. A common result is that in situ measurements are more accurate for very-short term forecasts (real-time to one hour ahead forecasts), satellite data is better for short-term forecasts (up to 6 hours ahead), and Numerical Weather Predictions (NWP) are better for long-term forecasts (day-ahead and longer). Models also vary, with auto-regressive approaches being commonly used for very-short term forecasts, while longer forecast horizons use a wide range of machine learning models.

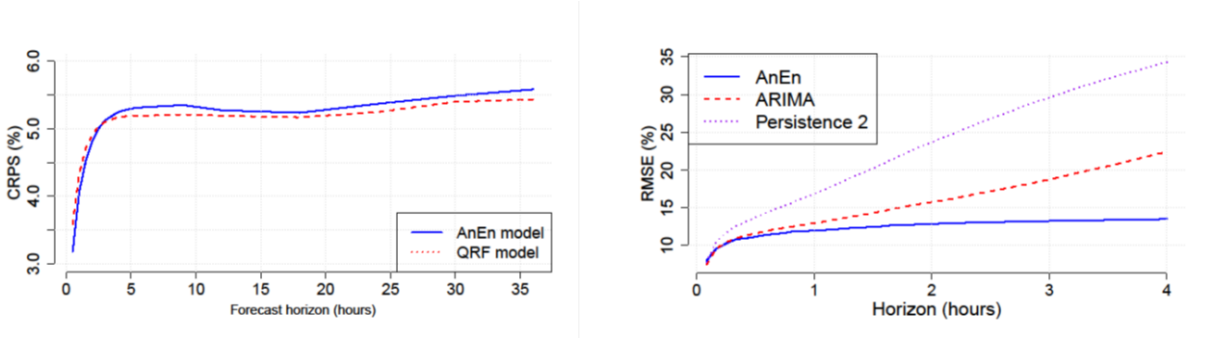
The RES producers have thus to develop and maintain numerous forecasting models for the different decision-making processes they are involved in, usually fitted for each power plant. This increases further the complexity of the decision-making processes and can create problems regarding the continuity of the forecasts.

In this work we propose a forecasting model for PV power generation that can use all the inputs mentioned before, and weights them according to the forecasting horizon. It can thus operate from very short-term to day-ahead forecast horizons with state-of-the-art performance. It can also directly provide probabilistic forecasts for an aggregation of power plants, thus allowing having a single forecasting model for managing a virtual power plant. The model follows the “lazy learning” paradigm, where generalization from the training set is only computed when a forecast is requested. Thus, the model is resilient to changes in the neighborhood of the plant (surrounding environment, partial outage, soiling, etc.)

The model is based on the Analog Ensemble (AnEn) method. However it is structurally expanded to allow the method to use an arbitrary large number of inputs. Each input is then weighted depending on the forecast horizon. As an example, for a given input for one-hour ahead, the weight is computed

based on the Mutual Information (MI) between the input and the PV power generation observed one hour later. This allows dynamically selecting the most relevant inputs depending on the horizon.

The model is evaluated for short-term and day-ahead forecasts, and compared with a Quantile Regression Forest (QRF) for day-ahead forecasts, and a linear Auto-Regressive Integrated Moving Average (ARIMA) model for the short term forecasts. Results show that the AnEn model is competitive with the QRF model in day-ahead forecasting. It is also consistently better than the ARIMA model for short-term forecasting.



Comparison of the AnEn model 30-min resolution forecasts for day-ahead, and 5-min resolution forecasts for short-term. The lower the CRPS or RMSE, the more accurate the model